

Approaches to Defining Topic-Related Portfolios of Biomedical Research

Danielle Dae, Elizabeth Hsu and James Corrigan



Abstract

The fundamental starting point for any portfolio analysis is the definition of the portfolio area. Biomedical research funders often describe and analyze portfolios of projects considered relevant to specific research areas, typically a disease, a group of diseases, research approach, or scientific topic. Setting boundaries on sets of topic-relevant projects can be complex, and approached multiple ways. Common approaches rely on manual expert review and/or automated text mining to classify project proposal documents. These common prospective approaches can be less than accurate, with potential for both under- and over-reporting errors. A retrospective approach that classifies research based on project outputs (i.e. publications) may improve portfolio estimates. We present a case study using the National Cancer Institute's (NCI's) pediatric cancer portfolio to compare two prospective approaches (manual classification and text mining-based classification) and a retrospective approach of classification through publication-based linkages. While there is overlap among the projects captured by the three approaches, each approach also identifies a unique subset of projects. Analysis of projects reveals that publication-based approaches can augment portfolios by capturing additional projects.

Approach

INDEXING
Prospective

EXPERT REVIEW
Prospective

PUBLISHED RESULTS
Retrospective

Topic Source

Research Proposal

Research Proposal

Research Outputs

Data Source

IMPAC II

NCI Funded Research Portfolio (NFRP)

PubMed Indexed Articles

Code of Interest

RCDC categories and terms
(assignment to expert-defined categories/terms is determined by text mining title, abstract, specific aims, and public health relevance section)

NCI-specific coding, Special Interest Categories
(assignment by expert review of entire research proposal)

keywords in published title or abstract

Strategy

- FY11 NCI projects associated with RCDC categories or terms and keywords relevant to pediatric research and cancer were identified using the Query, View, Report (QVR) system
- Pediatric categories/terms: Pediatric, Pediatric AIDS, Pediatric Cancer, Pediatric Research Initiative, Pediatric Oncology, Pediatric Neoplasm, or Malignant Childhood Neoplasm
- Cancer categories/keywords: Malignant Neoplasm, Cancer, Cancer

- FY11 NCI projects associated with the Special Interest Category "childhood cancer" were identified in NFRP (<http://fundedresearch.cancer.gov/ncipportfolio/>)
- Intramural projects, contracts, and general support awards were excluded since they are not reliably captured by other approaches

- Publications associated with NCI grant numbers were searched for the keywords cancer and either pediatric or childhood in PubMed.
- (["NCI"[Grant Number] OR "CA"[Grant Number] OR "CA"[Grant Number] OR "CA*"[Grant Number]]) AND (["pediatric"[Title/Abstract] OR "childhood"[Title/Abstract]) AND "cancer"[Title/Abstract])
- NCI grant numbers were extracted and matched to comprehensive FY11 NCI project lists to identify FY11 NCI projects that published pediatric relevant research

Portfolio

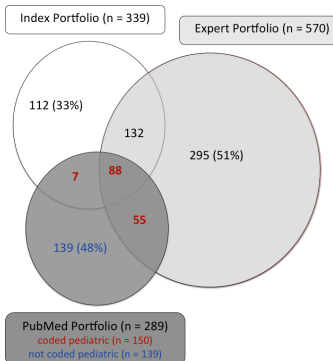
Index Portfolio, 339 unique projects

Expert Portfolio, 570 unique projects

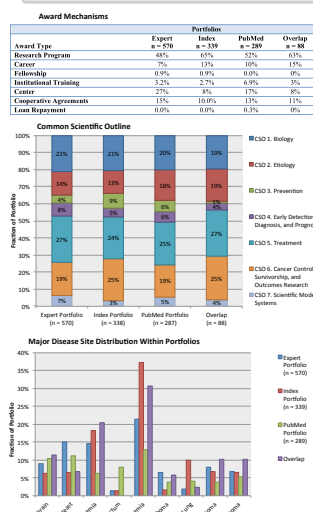
PubMed Portfolio, 289 unique projects

Results

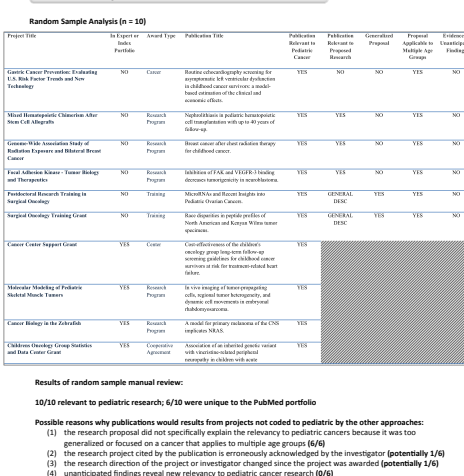
Mapping projects across portfolios



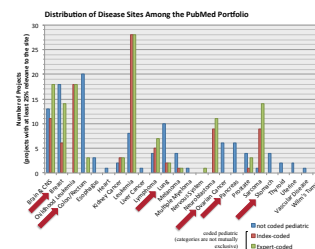
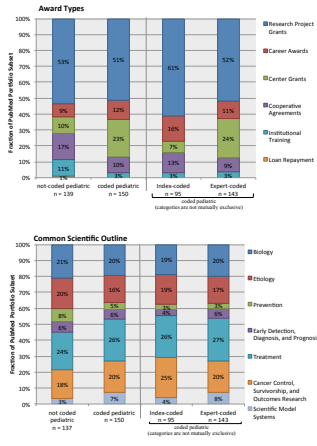
Comparisons across portfolios



Analysis of PubMed Portfolio



Comparing PubMed portfolio projects, coded vs. not-coded pediatric



Enriched sites manual review: Colon Test Case (n=20)

- 6/20 articles not relevant to pediatric cancer research because (1) the published research examined the effect of factors that spanned an individual's lifespan on adult cancers (5/6) or (2) pediatric cancer findings were mentioned for comparison, but research findings were relevant to adult cancers (1/6)
- 14/20 articles were relevant to pediatric cancer research projects
 - 6/14 projects were also relevant to pediatric-relevant sites (Brain and Leukemia)
 - 3/14 projects had generalized research proposals
 - 1/14 projects are focused on childhood cancer survivors
 - 3/14 apply a method relevant to their proposed research to study a pediatric-relevant topic
 - 1/14 projects is a reporting error (author project number transposition)

Conclusions

- Retrospective approaches can be used to enrich a portfolio with projects that publish relevant results despite proposing research that did not specifically explain the relevancy to the specified topic because it was too generalized or focused on a topic with broader implications
- Each approach to defining portfolios identifies a unique subset of projects suggesting that using multiple approaches allows one to converge on a more robust portfolio
- Future Directions: deeper analysis to understand why the different approaches capture unique projects

Abbreviations

RCDC: Research, Condition, and Disease Categorization
NFRP: NCI Funded Research Portfolio (<http://fundedresearch.cancer.gov/ncipportfolio/>)